Hadoop MapReduce for seismic detection using STA/LTA trigger algorithm

Youness Choubik[1], Abdelhak Mahmoudi[2], Mohammed Majid Himmi[3]
[1,2,3] LIMIARF Laboratory, Faculty of Sciences, Mohammed V University Rabat, Morocco
[2]Ecole Normale Supérieure, Mohammed V University Rabat, Morocco
[1]youness.choubik@gmail.com, [2]abdelhak.mahmoudi@gmail.com, [3]himmi.fsr@gmail.com

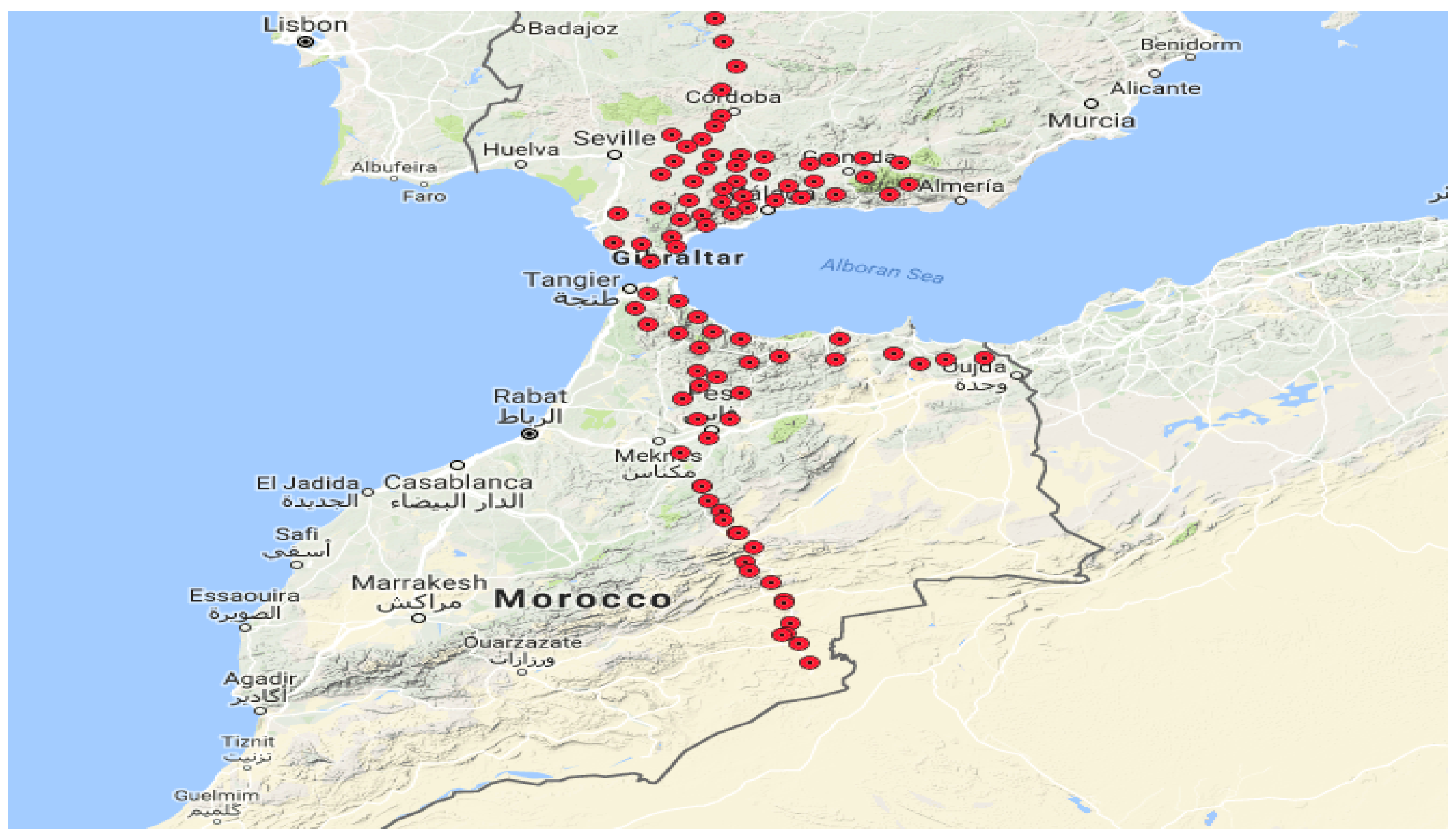
INTRODUCTION

In recent years, processing data using traditional tools has become a difficult task due to the huge amount of available data. Hence the need of new tools and frameworks that facilitate and accelerate data processing. Big Data tools have become widely used in many fields, including Seismology.For example, [1] cross correlated a global dataset consisting of over 300 million seismograms. By using a Hadoop cluster they achieved a factor of 19 performance increase on a test dataset. [2] used Hadoop and Spark to perform a large-scale calculation of seismic waveform quality metrics. [3] have developed a big data platform, which is built upon Apache Spark. [4] has adopted Hadoop platform to manage, store, and to analyze seismic data.In this work we used Hadoop MapReduce to implement STA/LTA trigger algorithm, which is widely used in seismic detection. This implementation allows to find out how effective it is in this type of tasks as well as to accelerate the detection process.

METHOD

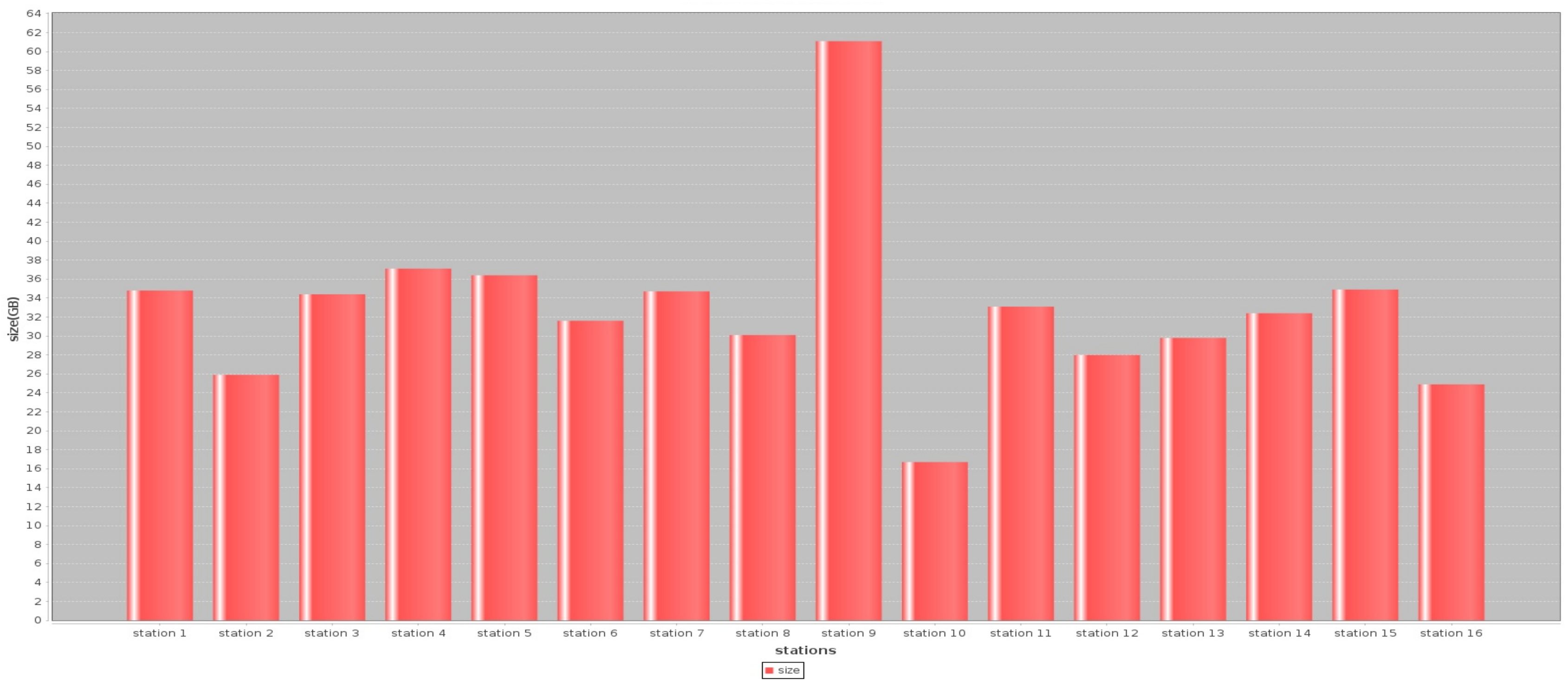
The dataset used in this work is from a seismological network called XB which was deployed in both Morocco and Spain, between the time period 2009 to 2013.Figure 1 shows the positions of the XB network.

Figure 1 – XB seismic network



The seismic data are SEED files. To test our implementation, we used a dataset of 14 stations from the XB seismological network. Figure 2 shows the size of the Avro files corresponding to every seismic station.

Figure 2 – Size of Avro files of the 14 PM Stations



We used Hadoop framework to process our dataset across a cluster of commodity hardware. The configurations of the cluster used in this work is given in table 1.

	master nodes	worker nodes
Count	1	5
Storage	100 GB	900 GB
Memory	5 GB	25 GB
Cores	2	5

Table 1 – Hadoop cluster's configuration

In order to find seismic events and the stations that simultaneously trigger them, we used two MapReduce functions. In the first map function, we defined two windows, one for calculating the Short Time Average (STA window) and the other for the Long Time Average (LTA window).The above averages are used to calculate the STA/LTA ratio. We consider an assumed seismic event in a single component when the ratio exceeds a defined trigger threshold.

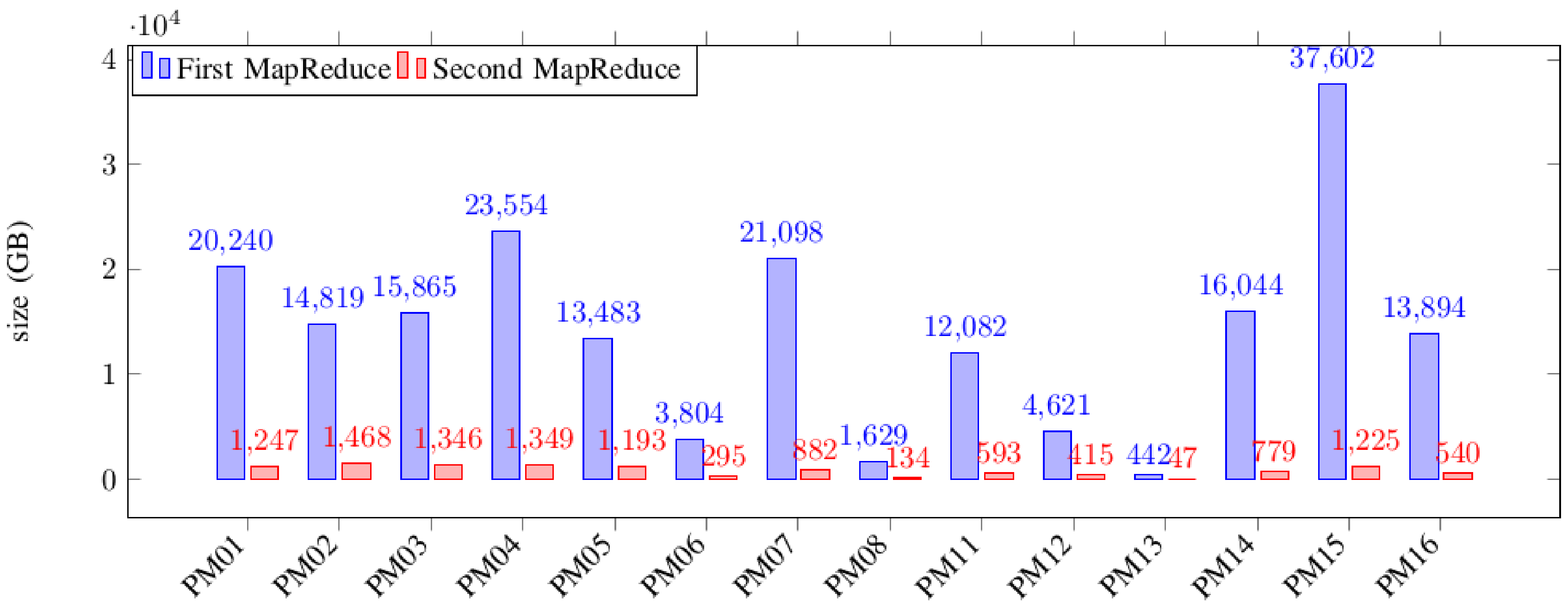
In the first Reduce function the list of value for each key(station) is sorted by the start time, then the whole list is treated so that if the event is detected in only one channel it is considered as a noise. Otherwise, if it is detected in more than one channel in the same station this event is considered as an assumed earthquake.

The second Map function just returns the data returned by first reduce function. In the second Reduce function the values returned by the second Map function are sorted by detection time and each event is compared with the events that follow. When the interval of time between two events is lower than a defined value, the function check if the stations where the events occurred are neighbors.

RESULTS

To find out the improvement realized by MapReduce we considered a traditional implementation as a reference and compared its results with that of MapReduce. The time needed for processing data by the reference implementation was almost 13 hours and half, while MapReduce needed nearly 9 hours to accomplish the tests. The MapReduce implementation decreased the processing time by 34%. We were able to detect 199177 events in the first MapReduce function. By applying the second MapReduce, the number of events detected in more than one station decreased to 11513 events. Figure 3 shows the number of events in each seismic station.

Figure 3 – Number of seismic events detected per Station



CONCLUSION

We have shown in this paper the usability of Hadoop MapReduce for seismic detection, particularly using Short Term Average to Long Term Average (STA/LTA) algorithm. We compared MapReduce implementation performance with that of a traditional implementation. The results show that time needed for processing goes from almost 13 hours and half using the traditional implementation to nearly 9 hours by using MapReduce. So MapReduce decreases the processing time needed for processing large amount of seismic data. Looking forward, our goal is to apply seismic detection on the entire XB network. This requires further optimizing since Short Term Average to Long Term Average lacks accuracy, and we should combine other techniques to improve its detection accuracy

REFERENCES

[1] T.G. Addair, D.A. Dodge, W.R. Walter, and S.D. Ruppert. Large-scale seismic signal analysis with hadoop. Comput. Geosci., 66(C):145–154, May 2014.

[2] S. Magana-Zook, J.M. Gaylord, D.R. Knapp, D.A. Dodge, and S.D. Ruppert. Large-scale seismic waveform quality metric calculation using hadoop. Computers & Geosciences, 94(Supplement C):18 – 30, 2016. ISSN 0098-3004.

[3] Lei Huang, Xishuang Dong, and T. Edward Clee. A scalable deep learning platform for identifying geologic features from seismic attributes. The Leading Edge, 36(3):249–256, 2017.

[4] Zhuang Chen and Ti Zhang. Processing and analysis of seismic data in hadoop platform, 02 2017.